



## COVER SHEET

---

**This is the author-version of article published as:**

Sullivan, Karen A (2005) Alternate forms of prose passages for the assessment of auditory-verbal memory. *Archives of Clinical Neuropsychology* 20(6):pp. 745-753.

**Accessed from** <http://eprints.qut.edu.au>

**Copyright**2005 Elsevier

**Note:** For access to the prose passages referred to in this paper, please contact the author [ka.sullivan@qut.edu.au](mailto:ka.sullivan@qut.edu.au)

Running Head: Alternate forms of prose passages

Alternate forms of prose passages for the assessment of auditory-verbal memory.

Karen Sullivan, PhD

Queensland University of Technology

AUSTRALIA

#### Acknowledgements

Some of this research was completed by Karen Sullivan as part of her PhD, under the supervision of Dr Stephen Bowden at the University of Melbourne, Australia.

The sample used for the second set of analyses reported in this paper has been described previously as have parts of the procedure (Sullivan, 1996).

The author would like to thank independent raters who participated in this project, and the Queensland University of Technology Library Staff, particularly Stephanie Bradbury for help with the proof-reading of this manuscript.

The writing up of this research was funded by a writer-in-residence grant awarded to Karen Sullivan from the Research and Postgraduate Committee of the School of Psychology and Counselling, Queensland University of Technology.

Correspondence concerning this article should be addressed to Karen Sullivan at the School of Psychology and Counselling, Queensland University of Technology, Carseldine, Queensland 4034 Australia. Electronic mail should be sent to [ka.sullivan@qut.edu.au](mailto:ka.sullivan@qut.edu.au)

*Key words:* Wechsler Memory Scale, free recall, test forms, verbal memory, psychological assessment

### Abstract

Logical Memory (LM) is the most frequently administered subtest from the Wechsler Memory Scale; however the lack of alternate equivalent forms for this subtest may limit its clinical utility. Six new paragraphs modelled on LM stories were developed. Stories were matched on attributes such as number of words and readability. Passage attributes for the six stories were compared with those of standard LM stories (WMS-R and WMS-III versions) to examine story equivalence. The psychometric properties of new passages were also calculated to assess task difficulty and interrater reliability. Results from these analyses suggest a high degree of overlap between the attributes of the new stories and some interesting discrepancies between passage attributes of WMS-R and WMS-III LM stories. In addition, interrater reliability of new passages was found to be excellent (at least .97), and when combined into three sets of passage-pairs, these pairs were found have equivalent difficulty. To reduce the potential for practice effects by use of alternate forms, these new logical memory-style passages may be particularly useful for repeat assessment of auditory-verbal memory.

### Alternate forms of prose passages for the assessment of auditory-verbal memory.

There appears to be a growing need to develop alternate forms of tests given recent claims that repeated administration of neuropsychological tests is now "commonplace" (Benedict & Zgaljardic, 1998). In recognition of this need, there have been several papers describing the development of alternate forms of a number of neuropsychological tasks and subtests.

Examples include the Wisconsin Card Sorting Test (Ferland, Ramsay, Engeland, & O'Hara, 1998), Boston Naming Test (Fastenau, Denburg, & Mauer, 1998), Auditory Verbal Learning Test (Uchiyama et al., 1995), Expanded Halstead-Reitan Neuropsychological Test (Dikmen, Heaton, Grant, & Temkin, 1999) and other non-memory verbal and spatial processing tasks (Watson, Pasteur, Healy & Hughes, 1994; Zgaljardic & Benedict, 2001). It should also be noted that whilst the existence of alternate forms for popular tests such as the Auditory Verbal Learning Test and LM is not new (see Crawford, Stewart, & Moore, 1989; Lezak, 1995), there appears to be increasing recognition of the need to empirically demonstrate equivalence across forms and characterize the psychometric properties of alternate forms of tests more fully than may have been the case previously.

LM is one of the tests that has been the focus of efforts to develop psychometrically equivalent forms (e.g., Morris, Kunka & Rossini, 1997). Given that this test is one of the most frequently used subtests from the Wechsler Memory Scale (WMS) in North America (Butler et al., 1991) and Australia (Sullivan & Bowden, 1997) and it has relatively good psychometric properties (Spren & Strauss, 1998), this focus is perhaps not surprising. In addition, the susceptibility of LM to substantial practice effects has been well documented (e.g., Benedict & Zgaljardic, 1998), further highlighting the need for alternate forms of this test.

The "Morris revision" of LM consists of two alternate paragraphs designed to be psychometrically equivalent to the WMS-R stories (Morris et al., 1997). These paragraphs

feature Greg and Martha respectively. Morris and colleagues attempted to match their stories with the WMS-R stories on a number of passage attributes, such as the number of scorable ideas, subjective passage attributes (specifically novelty and affective tone), readability, interscorer reliability, and correlation between story pairs. The results of comparisons conducted by Morris et al. (1997) suggest these authors generally succeeded in generating a pair of stories that match the WMS-R Logical Memory (LM-R) passages. For example, no significant differences were found on sets of stories administered in a serial counterbalanced fashion (ie. Anna + Robert followed by Greg + Martha, or vice versa). Second, significant positive correlations were found between sets of stories indicating a moderate-to-strong association between test sets.

The Morris et al.(1997) analysis also revealed some significant discrepancies on selected passage attributes *between* LM-R stories (e.g., The Anna Thompson story has fewer (half as many) longer sentences and a more complex grammatical structure than the Robert Miller story). The Morris stories reflect the inequities of the WMS-R passages in terms of such attributes (e.g., the Greg story also has fewer<sup>1</sup> longer sentences than the Martha story).

The development of the Morris stories was reported as a pilot study (Morris et al., 1997, p. 371) and it has helped to highlight the need for alternate forms of the LM stories. However the Morris stories could only be matched to the LM-R stories since the WMS-III LM stories had yet to be published (Wechsler, 1997a; 1997b)). Whilst it must be acknowledged that some clinicians may still be using the WMS-R (Tulsky & Ledbetter, 2000) and for such clinicians the Morris revision may be continue to provide a useful alternate form of LM, the WMS-III is generally regarded as a significant improvement over earlier versions of this scale (Axelrod, 2001; Ryan, Ament, & Arb, 2000) and alternate versions of WMS-III stories are needed.

---

<sup>1</sup> Half as many

Further, although the early work of Morris et al. was important because it provided neuropsychologists with alternate LM-R stories, a suite of more than two alternate stories may be needed. For example, neuropsychologists needing to track recovery of function throughout the course of a patient's stay in rehabilitation or over the first year or so following a head injury may need to conduct serial assessments more than twice.

Finally, although the Morris passages may be considered a significant first step towards the development of alternate Logical Memory stories, the content and structure of these passages deserves comment. In terms of content, the Morris stories include colloquialisms (e.g., the Greg story refers to "quarterbacks", and the Martha story refers to "bumble bees"). These inclusions may increase the likelihood that users of the Morris revision outside North America will modify story content as has been shown to be the case with LM-R stories (Sullivan & Bowden, 1996). Such modifications have the potential to increase the likelihood of unstandardized administration and need to be avoided in future attempts to devise alternate LM stories. In terms of structure, there are some criteria on which the Morris stories do not appear to be well matched (for example the "Greg" story is 10 words (one sixth) shorter than the "Martha" passage). These structural variations occur in addition to the inequities between LM-R passages that are reflected in the Morris passages, and future attempts to devise alternate LM stories might be wise to eliminate such variation.

The aim of this study was to devise a more extensive set of LM-style prose recall passages that would be free from colloquialisms and useful for repeat administration. Two sets of analyses were carried out as part of the process used to develop these passages. The first set of analyses was aimed at investigating the equivalence of alternate LM stories using criteria similar to those used by Morris et al. (1997). Further, to explore comparability with existing Logical Memory stories these analyses included data from standard WMS-R and

WMS-III Logical Memory stories. The second set of analyses was intended to explore the psychometric properties of the six new alternate prose recall stories.

## Method

### *Materials*

Eight new passages (pilot versions) were devised with approximately the same number of words, and scoring units as LM-R stories<sup>2</sup>. The content of stories was based on short newspaper items and was designed to be similar to WMS-III passages in terms of affective tone and style. To eliminate the inclusion of colloquialisms, comments were sought from five reviewers with training in neuropsychology on pilot versions of alternate passages. Pilot reviewers were instructed to pay particular attention to the identification of colloquial phrases and following identification these were removed from the final versions of stories. Six stories of the eight pilot stories were retained following review. These stories are referred to in the following sections of text by a key descriptive word from each story (e.g., Tokoyo or trolleys etc).

Scoring guidelines similar to those used by Wechsler (1987; 1997a; 1997b) were devised for each passage and a record form constructed based on Wechsler's (1987; 1997a; 1997b). Full scoring guidelines and a copy of the record form devised for this study are available on request<sup>3</sup>.

### *Analysis I: Passage characteristics*

Table 1 includes an analysis of passage characteristics for WMS-R (Anna, Robert), WMS-III (Anna, Joe), and alternate passages devised for this study. There are some interesting trends apparent in this Table. For example consistent with Morris et al.'s findings, Table 1 shows there are considerable differences between LM-R passages in the

---

<sup>2</sup> Note that the number of words on WMS-III stories differs substantially (see Table 1), and as such this version of the stories were considered less optimal as a model for new passages. The number of words in WMS-R passages is approximately 66 and new passages were designed to match with this standard. Both WMS-R and WMS-III stories include 25 scoring units as do the new passages.

number of sentences per paragraph and the readability of stories (i.e., the Anna Thompson story contains fewer longer sentences and requires more advanced comprehension when this is defined in terms of readability statistics than the Robert Miller story). The new Joe Garcia story (which replaces Robert Miller on the WMS-III) has 30% more words than Anna Thompson. Importantly, the alternate passages devised for this study show a reasonable degree of consistency with each other on all of the parameters shown in Table 1 and are more like the Robert Miller or Joe Garcia stories, in that they have more, shorter sentences than Anna Thompson, and consequently better readability.

---

Insert Table 1 about here

---

To provide an index of variation among stories on the first eight parameters listed in Table 1, the mean and standard deviation for each index was calculated across nine passages. A passage was defined as extreme on a criterion if an item was more than two standard deviations above or below the mean for that index. For example, Table 1 shows that the WMS-III Joe Garcia passage has significantly more words (84) than the other eight passages ( $M = 68$ ). Similarly, the Anna Thompson story has significantly fewer sentences per paragraph (1.5) than other passages ( $M = 4$ ). The other WMS-R passage, Robert Miller, differs from the other eight stories on three of the first eight parameters shown in Table 1. Specifically, the Robert Miller story has significantly more sentences, fewer words per sentence and is easier to read (based on Flesch-Kincaid grade level scores) than other passages.

Although, the relationship between parameters listed in Table 1 and the degree of memorability of passages has yet to be fully determined, a lack of comparability on indices may have implications for task equivalence. For example, task equivalence of LM-R stories has been studied by several investigators (e.g., Ivison, 1993; Petrick, Haut, & Franzen, 1993;

---

<sup>3</sup> Materials were reproduced in text because of space constraints.



Rankin & Gore, 1994). Results from comparisons of LM-R stories suggest that stories A and B are equivalent tasks using standard- (Rankin & Gore, 1994), or Australian-versions of WMS-R stories (Iverson, 1993). However, task equivalence between LM-R stories should be regarded as a tentative finding given that other studies (e.g., Petrick et al., 1993) have suggested that LM-R story A (Anna Thompson) is more difficult than LM-R story B (Robert Miller). Arguably, the variability among studies on LM-R story equivalence may be partly due to the differential passage characteristics of LM-R stories, such as differences in the readability and grammatical complexity of these passages.

*Analysis II: Investigation of the psychometric properties of alternate stories*

*Participants*

Thirty-two undergraduate psychology students participated in this part of the project. These participants were part of a larger study which has been reported previously (see Sullivan, 1996). The majority of participants in this sample were female (84% of the sample) and the average age of participants was 21 years ( $SD = .56$ ).

*Procedure*

Subjects were individually administered eight prose recall stories (six experimental alternate forms plus two WMS-R Logical Memory stories) by one of two trained psychologists. To ensure standard passage presentation, stories were pre-recorded on audio-tape and played to subjects during testing given that presentation rate has been shown to have a significant effect on logical memory recall (Shum, Murray, & Eadie, 1997). Using Wechsler's standard instructions for LM, subjects were asked to listen to and try to remember each story, and then tell it back. Passages were presented in a counterbalanced order using a latin-squares design, such that each story was presented in one of eight possible positions as often as every other story.

To supplement standardized recording procedures, minimize examiner error, and allow for calculation of inter-rater reliability statistics (also see Sullivan, 1996), responses were recorded on audiotape. This procedure was used in addition to the standard procedure of checking off or marking responses directly on the test record form (see Spreen & Strauss, 1998). Note that the use of taped LM responses has been incorporated in the WMS-III as an optional scoring method for this version of Logical Memory. Taped responses were copied and sent to three independent trained raters for scoring, together with newly devised record forms for the six passages. Raters were selected on the basis that they were experienced clinical neuropsychologists trained in the administration of LM.

### Results

Descriptive statistics for each of the six new passages devised for this study were calculated, together with estimates of interscorer reliability. These figures are shown in Table 1. This Table shows that the mean score for passages ranged from approximately 13 to 16 out of 25, with relatively small standard deviations (i.e., between three and four points). Data on the interscorer reliability of passages calculated across three independent raters suggests good interrater reliability. In addition, it should be noted that the interscorer reliability for the six new stories is comparable to interscorer correlation of .97 reported for LM-R passages (Wechsler, 1987) and the WMS-III<sup>4</sup>.

Data from stories and raters was further analysed using a 3 X 6 repeated measures ANOVA with one within-subjects factors (stories) and one between-subjects factors (raters). Results revealed a non-significant interaction between raters and stories,  $F(10, 178) = .179, p = .998$ , but a significant main effect for stories,  $F(5, 89) = 19.583, p = .000$ . Further analysis of this significant main effect using simple contrasts suggests this difference was due to differential performance on the Simpson and Trolley stories (higher scores on the former;

lower scores on the latter) relative to the Tokyo story. However the magnitude of this difference appears to be relatively small (i.e., two scorable ideas out of 25; see Table 1) and may not be clinically meaningful. This suggests that of the new passages devised for this study, four out of six passages were of equivalent difficulty, with minor variation on the Simpson and Trolleys stories respectively. When these two stories were paired together to enable analysis of scores on story sets (or pairs), this difference was negated (see below).

*Descriptive statistics for story pairs.* Logical Memory scores are typically based on scores that reflect recall across two stories (e.g., Wechsler, 1987; 1997b)<sup>5</sup>, therefore descriptive statistics for three story pairs based on the six new passages were also calculated. The mean score on story pairs was approximately 28 when the following stories were combined: Europe and Tokyo ( $M = 28.27$ ,  $SD = 6.29$ ), Simpson and Trolleys ( $M = 28.52$ ;  $SD = 7.40$ ), and Granger and Yorkshire ( $M = 28.92$ ,  $SD = 6.42$ ). A repeated measures ANOVA with one within-subjects factor (story pair) revealed no significant difference between the mean score on each of the three story pairs,  $F(2, 94) = 1.039$ ,  $p = .358$ . According to the WMS-R manual (Wechsler, 1987) the mean score for LM immediate recall) is 25.7 (20-24 year-old age group). Comparable data is not reported in the WMS-III technical manual (Wechsler, 1997a; 1997b) since the scoring of the WMS-III version of LM now includes the second recall score from story B. In summary, the mean score for the six new passages devised for this study is of a similar order of magnitude to that reported for standard WMS-R stories and the Morris Revision ( $M = 25.10$ ).

*Correlations between story pairs.* The intercorrelation matrix between stories comprising pairs was generated to investigate the relationship between elements. Total story

---

<sup>4</sup> Interscorer reliability coefficients for the Logical Memory I subtest of the WMS-III are described as “greater than .90” Wechsler, 1997b, p.65)

<sup>5</sup> WMS-R LM total recall scores are based on the sum of units recalled from story A plus story B (Wechsler, 1987). WMS-III LM total recall scores are based on the sum of units recalled from stories A and B, *plus* the number of units recalled from a second recall of story B (Wechsler, 1997a).

recall scores averaged across three raters were used for this analysis. Consistent with expectations, story-pairs were significantly correlated (i.e., Europe-Tokyo pair,  $r = .409$ ; Simpson-Trolleys pair,  $r = .423$ ; Granger-Yorkshire pair,  $r = .512$ ). These results suggest that scores on stories combined in pairs are moderately-to-strongly positively correlated.

### Discussion

The aim of this study was to develop alternate, equivalent prose-recall passages, modelled on the Logical Memory stories of the Wechsler Memory Scales. Use of alternate forms of verbal memory tests has been shown to be particularly important in reducing the magnitude of practice effects when repeat testing (Benedict & Zgaljardic, 1998; Zgaljardic & Benedict, 2001). Previously, only one alternate form of the WMS-R LM stories was available, with no published alternatives to the WMS-III stories.

To achieve this aim, six new LM-style passages were devised in a two-stage process. The first stage involved development and pilot testing of passages, with specific analysis of passage characteristics. The second stage involved investigation of the psychometric properties of new passages. The results of this study suggest that the passages developed for this study have similar attributes and comparable psychometric properties both with each other, and with the WMS-R stories. Further research is needed to empirically test the equivalence of WMS-III stories and the new passages devised for this study.

Although efforts to develop alternate forms of Logical Memory passages are not new (e.g., Morris et al., 1997), the stories devised in this study are unlike those devised previously in that they do not mirror inequities in passage characteristics of WMS-R stories (or indeed in WMS-III Logical Memory stories). Passages devised in this study also differ from existing alternate Logical Memory stories in that the stories devised for this study were designed to exclude colloquial content as a way of increasing the likelihood of standard administration.

In terms of the clinical application of the passages described in this study, caution is warranted. Whilst Morris et al. (1997) advised that normative data for standard LM-R passages could be used to interpret performance on their stories, this may not be advisable with the stories described here since they do not share the same degree of similarity with LM-R stories as the Morris revision. Similar arguments can be made about the use of WMS-III normative data to interpret performance on the stories described here. In addition, although this study provides preliminary data on how a select group of non-clinical participants perform on these passages, the extent to which these results generalize to non-clinical samples has yet to be demonstrated. Extensive investigation of the normative properties of these stories needs to be undertaken before these passages are used for non-experimental purposes in clinical settings. The exception to this could include clinical applications that involve ipsative comparisons where data provided by individuals on one occasion serves as a reference point for subsequent evaluations and comparisons. Finally, notwithstanding these interpretative caveats, preliminary investigations of the passage attributes and psychometric properties of the passages reported here suggest these stories will prove useful in a variety of clinical and research settings, particularly those in which repeat assessment of auditory-verbal memory using alternate forms needs to be undertaken.

## References

- Axelrod, B. N. (2001). Administration duration for the Wechsler Adult Intelligence Scale-III and the Wechsler Memory Scale-III. *Archives of Clinical Neuropsychology*, 16, 293-301.
- Benedict, R. H. B., & Zgaljardic, D. J. (1998). Practice effects during repeated administration of memory tests with and without alternate forms. *Journal of Clinical and Experimental Neuropsychology*, 20, 339-352.
- Butler, M., Retzlaff, P., & Vanderploeg, R. (1991). Neuropsychological test usage. *Professional Psychology: Research and Practice*, 22, 510-512.
- Crawford, J.R., Stewart, L.E., & Moore, W. (1989). Demonstration of savings on the AVLT and development of a parallel form. *Journal of Clinical and Experimental Neuropsychology*, 11, 975-981.
- Dikmen, S. S., Heaton, R. K., Grant, I., & Temkin, N. R. (1999). Test-retest reliability and practice effects of Expanded Halstead-Reitan Neuropsychological Test Battery. *Journal of the International Neuropsychological Society*, 5(4), 346-356.
- Fastenau, P. S., Denburg, N. L., & Mauer, B. A. (1998). Parallel short forms for the Boston Naming Test: Psychometric properties and norms for older adults. *Journal of Clinical and Experimental Neuropsychology*, 20, 828-834.
- Ferland, M. B., Ramsay, J., Engeland, C., & O'Hara, P. (1998). Comparison of the performance of normal individuals and survivors of traumatic brain injury on repeat administrations of the Wisconsin Card Sorting Test. *Journal of Clinical and Experimental Neuropsychology*, 20(4), 473-482.

Iverson, D. (1993). Logical memory in the Wechsler Memory Scales: Does the order of passages affect difficulty in a university sample? *Clinical Neuropsychologist*, 7(2), 215-218.

Lezak, M. D. (1995). *Neuropsychological assessment* (3rd ed.). New York: Oxford University Press.

Morris, J., Kunka, J. M., & Rossini, E. D. (1997). Development of alternate paragraphs for the Logical Memory subtest of the Wechsler Memory Scale-Revised. *The Clinical Neuropsychologist*, 11, 370-374.

Petrick, J.D., Haut, M. W., & Franzen, M.D. (1993). *The equivalence of Logical Memory stories of the WMS-R*. Paper presented at the 101st Annual American Psychological Association Convention, Toronto, Canada.

Rankin, E. J., & Gore, P. A. (1994). Task equivalence and a failure to detect interference effects among the logical memory stories of the Wechsler Memory Scale Revised. *Perceptual and Motor Skills*, 78, 291-294.

Ryan, J. J., Ament, P. A., & Arb, J. D. (2000). Supplementary WMS-III tables for determining primary subtest strengths and weaknesses. *Psychological Assessment*, 12, 193-196.

Shum, D. H. K., Murray, R. A., & Eadie, K. (1997). Effect of speed of presentation on administration of the Logical Memory subtest of the Wechsler Memory Scale-Revised. *The Clinical Neuropsychologist*, 11, 1-4.

Spren, O. & Strauss, E. (1998). A compendium of neuropsychological tests (2nd ed.). New York, NY: Oxford University Press.

Sullivan, K. A. (1996). Estimates of interrater reliability for the Logical memory subtest of the Wechsler Memory Scale-Revised. *Journal of Clinical and Experimental Neuropsychology*, 18, 707-712.

- Sullivan, K. A., & Bowden, S. C. (1996). Using Logical Memory in Australia. *Psychological Reports, 79*, 799-808.
- Sullivan, K. A., & Bowden, S. C. (1997). Which tests do Neuropsychologists use? *Journal of Clinical Psychology, 53*, 657-661.
- Tulsky, D. S., & Ledbetter, M. F. (2000). Updating to the WAIS-III and WMS-III: Considerations for research and clinical practice. *Psychological Assessment, 12*, 253-262.
- Uchiyama, C. L., D'Elia, L. F., Dellinger, A. M., Becker, J. T., Selnes, O. A., Wesch, J. E., Chen, B. B., Satz, P., van Gorp, W., & Miller, E. N. (1995). Alternate forms of the Auditory-Verbal Learning Test: Issues of test comparability, longitudinal reliability, and moderating demographic variables. *Archives of Clinical Neuropsychology, 10*(2), 133-145.
- Watson, F. L., Pasteur, M. A. L., Healy, D. T., & Hughes, E. (1994). Nine parallel versions of four memory tests: An assessment of form equivalence and the effects of practice on performance. *Human Psychopharmacology, 9*, 51-61.
- Wechsler, D. (1987). *Manual for the Wechsler Memory Scale - Revised*. San Antonio, TX: Psychological Corp.
- Wechsler, D. (1997a). *Administration and scoring manual for the Wechsler Memory Scale - Third Edition*. San Antonio, TX: Psychological Corp.
- Wechsler, D. (1997b). *Technical manual for the Wechsler Adult Intelligence Scale - Third Edition and the Wechsler Memory Scale - Third Edition*. San Antonio, TX: Psychological Corp.
- Zgaljardic, D. J., & Benedict, R. H. (2001). Evaluation of practice effects in language and spatial processing test performance. *Applied Neuropsychology, 8*(4), 218-223.



Table 1.

*Passage characteristics for Logical Memory stories from WMS-III (Joe and Anna) and WMS-R (Anna and Robert), and six alternate form passages devised for this study (Europe to Tokyo). Passage characteristics are grouped into five categories representing counts (e.g., number of words), averages (e.g., average of sentences per paragraph), readability indices (e.g., Flesch score), performance attributes for this sample (e.g., mean score and standard deviation), and interscorer reliability statistics (i.e., Pearson correlations for each story, calculated across three independent raters).*

Index	Passage								
	Joe	Anna**	Robert	Europe	Granger	Simpson	Trolleys	Yorkshire	Tokyo
<b>Count</b>									
Words	84*	65	67	64	66	66	68	68	65
Chars	370	274	316	332	329	308	352	316	282
Sentences	5*	3	6*	4	4	4	4	4	4
<b>Averages</b>									
Sentences/para	5	1.5*	6	4	4	4	4	4	4
Words/sentence	16.8	21.6	11.1*	16	16.5	16.5	17	17	16.2
Characters/ word	4.3	4	4.5	5.1	4.8	4.5	5.1	4.5	4.2
<b>Readability</b>									
Flesch score	65.9	71.6	76.8	55.7	58	70.8	58.9	61.4	71.9
Flesch-Kincaid	8.2	8.6	5.3*	7.4	9.2	7.4	9.2	8.9	7.2
Grade level									
<b>Performance attributes</b>									
Mean score	-	-	-	13.64	14.03	16.23	12.65	14.55	14.48
Standard deviation	-	-	-	3.44	4.12	3.84	3.62	3.84	4.36
<b>Interscorer reliability</b>									
Pearson correlations	-	-	-	.97	.98	.98	.97	.98	.98

\* indicates item is more than two standard deviations above or below mean across nine passages.

\*\* The same version of the Anna Thompson story is used on the WMS-R and WMS-III.

